經濟時系列における回歸分析と變數誤差

阿部, 統

經濟論叢

第七十八巻　第一號

〔昭和三十一年七月〕

京都大學經濟學會

經濟時系列における回歸分析と變數誤差

阿部　統

一

＜爲替レートの切下げは、貿易收支の均衡回復に効果があるか＞という周知の爲替安定性論の命題を實地にテストするのに、多くの研究は、諸國の輸入需要の價格彈力性は一よりはるかに小さいがゆえに爲替切下げの效果はほとんど期待できないという。しかし、たいていのばあいはそのかげに、「これは短期分析である。したがって長期になれば、所得效果やその他がはたらいて、あるていどは貿易收支の改善が見られるだろう」と期待しているように見える。[1] これにたいして、現實に政策として爲替切下げが論ぜられるばあいには、しばしば、「爲替切下げは一時的には效果があるかもしれないが、やがて國內のインフレがその效果を相殺してしまうだろう」と、さいきんのメキシコの例などがもちだされる。この兩者の見解の相異は、國際經濟理論における古典派と近代派の對立にもむすびついてはなはだ興味ふかいが、[2] そのさい、妥協點のひとつとして雙方の立場から是認されていることは、價格彈力性の計算の方法および統計資料が不適切で信頼しがたいものであり、そのために計算結果がゼロの方向への偏りをもつというオーカットの反證である。[3] この反證は、一般に時系列回歸分析における統計誤差の處理の仕方の本

質にふれる問題なので、この小論において組織的にかんがえてみようと思う。

(1) たとえば、L. A. Metzler, "The Theory of International Trade," *A Survey of Contemporary Economics*, Vol. I. H. S. Ellis, ed. 1948, P. 245. 小島清譯「國際貿易の理論」（都留重人課監修「現代經濟學の展望」理論編I、一〇六頁）

(2) *International Economics—Studies*, 1951 および *International Trade and Economic Development*, 1953 の序説における J. Viner の見解、それにたいする G. Haberler の反論を見よ。G. Haberler, "The Relevance of the Classical Theory under Modern Conditions," *American Economic Review*, May 1954.

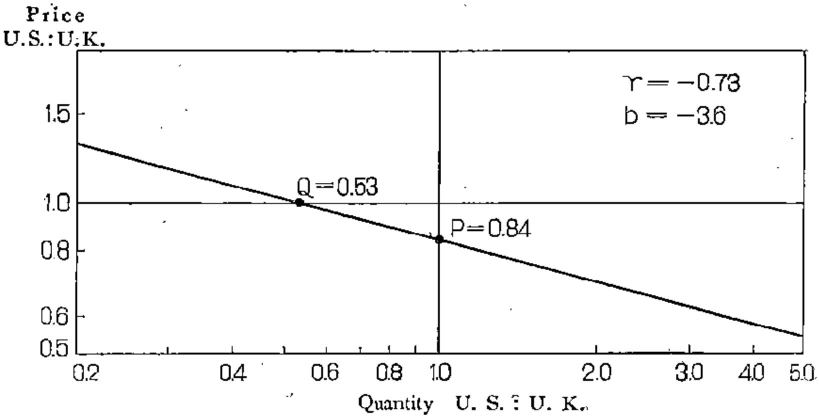
(3) G. H. Orcutt, "Measurement of Price Elasticities in International Trade," *Review of Economics and Statistics*, May 1950.

二

はじめに例をあげよう。マクドゥガルは比較生産費原理の妥當性を統計的に檢證する目的をもって、〈アメリカ（イギリス）の方がイギリス（アメリカ）にくらべて輸出量の多い製品は、はたしてアメリカ（イギリス）の方が比較優位にたっているかどうか〉を、まず〈アメリカ（イギリス）の方が相對價格が安いかどうか〉というかたちで、しらべようとした。[4] そのために、兩國からともに輸出される製品一〇九品目をえらび、一九三四—三八年の資料にもとづいて、それぞれの製品につき相對輸出量 $\frac{\text{Total quantity of U. S. exports, 1934-38}}{\text{Total quantity of U. K. exports, 1934-38}}$ および相對價格 $\frac{\text{Average value of U. S. exports, 1934-38}}{\text{Average value of U. K. exports, 1934-38}}$ を計算し、その間の回歸直線をもとめて第1圖のような結果をえた。[5] 相關係數は−0.74、スロープは−3.6である。つまり、資料によると、「一九三四—三八年の期間では、製品間に平均一パーセントの相對價格の差異があると、三・六パーセントの相對輸出量の差異をともなっていた」わけだが、マクドゥガルはこの回歸直線がもっと横にねる可能性があることを論じ、「統計資料に誤差がなければ、一パーセン

第1圖

Rerative Price & Quantities of U. S. & U. K. Exports

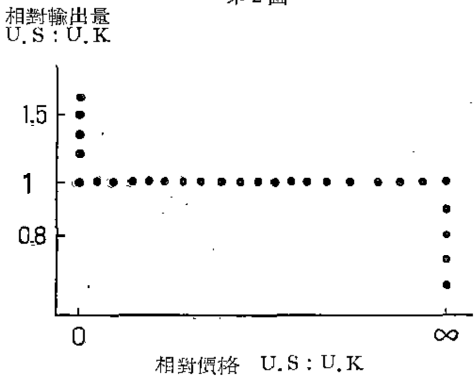


トの相對價格差は、すくなくとも四—四・五パーセントの相對輸出量差をともなうはず」として、つぎの論據を指摘する。

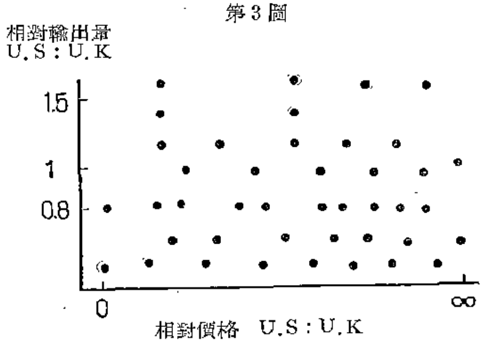
いま輸送費を無視すると、國際市場に完全競爭がおこなわれ、比較生産費原理が完全に妥當したばあい、統計資料に誤差がなければ兩國の相對價格と相對輸出量のあいだの關係は第二圖のようになるはずである。すなわち、アメリカの價格がイギリスの價格よりもすこしでも安いとイギリスからの輸出はまったくなくなり、イギリスの價格がアメリカの價格よりもすこしでも安いとアメリカからの輸出はまったくなくなる。しかしこのばあいでも、統計資料に觀察誤差がふくまれると、多かれ少なかれ觀測點は第三圖のようにちらばり、水平の方向に偏差の平方和を最小にしようとする回歸直線は傾斜をもたざるをえなくなる。いいかえれば、回歸方程式における獨立變數が觀察誤差をともなうとき、回歸係數の推定値をふつうの方法で計算するならば、その値は一般に眞の回歸係數よりも小になるのである。

この關係を數式によってしめせば、つぎのようになる。いま、相對價格の觀測値をX、相對輸出量の觀測値をYとし、各觀測値

第2圖



第3圖



はそれぞれ眞の値（X'・Y'）と觀察誤差（X''・Y''）の和であるとすれば

$$X=X'+X'' \qquad Y=Y'+Y'' \tag{1}$$

となる。ところでX'とY'のあいだに

$$Y'=\alpha+\beta X' \tag{2}$$

という關係があるものと想定すれば、(1)を(2)に代入して

$$Y=\alpha+\beta X+Y''-\beta X'' \tag{3}$$

をうる。$X \cdot Y \cdot X'' \cdot Y''$の平均値をそれぞれ$\bar{X} \cdot \bar{Y} \cdot \bar{X}'' \cdot \bar{Y}''$とし、$x \cdot y \cdot x'' \cdot y''$を各平均値からの偏差とすれば

$$X=\bar{X}+x \qquad Y=\bar{Y}+y$$

$$X''=\bar{X}''+x'' \qquad Y''=\bar{Y}''+y'' \tag{4}$$

となるから、(4)を(3)に代入して

$$Y=\beta x+y''-\beta x''+(\alpha-\bar{Y}+\beta\bar{X}+\bar{Y}''-\beta\bar{X}'') \tag{5}$$

をうる。右邊のカッコのなかはゼロである。(5)式に邊々xをかけて、あらゆるxについて總計すれば

$$\sum xy=\beta\sum x^2+\sum xy''-\beta\sum xx'' \tag{6}$$

したがって

$$\beta=\frac{\sum xy-\sum xy''}{\sum x^2-\sum xx''} \tag{7}$$

これが眞の回歸係數である。

しかし、統計資料をナマのままでもちいて、ふつうの方法で回歸係數を計算すると、推定値として

$$b=\frac{\sum xy}{\sum x^2} \tag{8}$$

をうる。βとbの大いさを比較するために

(i)各變數の眞の値と觀察誤差とのあいだには相關關係がない。すなわち

$$E(x'x'')=0,\quad E(y'y'')=0$$

(ii)從屬變數の觀察誤差は、獨立變數の眞の値と相關關係がない。すなわち

$$E(x'y'')=0$$

(iii) 兩變數の觀察誤差相互間には相關關係がない。すなわち

$$E(x''y'')=0$$

と假定して、(7)式を展開すれば

$$\beta=\frac{\sum xy-\sum x'y''-\sum x''y''}{\sum x'^2-\sum x'x''-\sum x''^2}$$

$$=\frac{\sum xy}{\sum x^2-\sum x''^2} \qquad (9)$$

となるから

$$\frac{b}{\beta}=\frac{\sum x^2-\sum x''^2}{\sum x^2}$$

$$=1-\frac{\sum x''^2}{\sum x^2}<1 \qquad (10)$$

より、$b<\beta$ をうる。つまり、マクドゥガルの論點が證明されたわけである。

(4) G. D. A. MacDougall, "British and American Exports: A Study Suggested by the Theory of Comparative Costs," *Economic Journal*, Dec. 1951.

(5) 圖は對數目盛をもちいてある。

(6) 森田優三「經濟變動の統計分析法」岩波全書。一六八—一六九頁。以下の證明も、本質的には森田教授の證明方法とかわらない。

三

さて、われわれが問題とするのは、統計資料が誤差をふくむばあい、上にのべたような回歸係數の推定値の偏りをどのようにして處理するか、ということである。この問題はすでにある程度古典的なものになっており、さまざまな解決への工夫がみられるが、いずれも經濟時系列に適用するにあたっては、多くの難點をのこしているように思われる。また、ふつうの實證的研究では、問題をまったく無視してしまって、單に古典的な回歸係數を計算するだけで安んじているが、このときにはその回歸係數が不偏推定値であるための假設を暗然のうちにみとめねばならない。このような假設が經濟現象にそのまま是認されるかどうかはきわめてうたがわしい。

ここで、以下の議論に卽して、もう一度問題の所在を定式化しておこう。簡單のために二變數 x・y 間の回歸關係を考え、常數項をゼロとする。變數の觀察誤差のほかに方程式誤差 ε を導入し、從屬變數が獨立變數以外の要因からうける disturbance をしめすものとすれば、

$$y'_i = a x'_i + \varepsilon_i$$

$$y_i = y'_i + y''_i \qquad x_i = x'_i + x''_i \qquad (i=1,2\ldots\ldots, n)$$

が成立する。ここに、x'・y'は各變數の眞の値を、x''・y''は觀察誤差をしめすものとする。

さて、x''・y''・εについてつぎの假設をもうける。

(i) $E(y''_i \mid x'_1, \ldots\ldots, x'_n\,;\, x''_1, \ldots\ldots, x''_n) = 0$

(ii) $E(x''_i \mid x'_1, \ldots\ldots, x'_n) = 0$

(iii) $E(\varepsilon_i \mid x'_1, \ldots\ldots, x'_n\,;\, x''_1, \ldots\ldots, x''_n) = 0$　　いずれも $i=1, 2, \ldots\ldots, n$

(i)は從屬變數の觀察誤差が獨立變數の眞の値に關しても、觀察誤差に關しても、獨立であることをしめし、(ii)は獨

立變數の眞の値と觀察誤差が相互に無關係であることをしめしている。また(iii)は方程式誤差が獨立變數と相關關係をもたないことを規定している。このばあい、各誤差の確率分布の平均値はすべてゼロとする。[9]

ところで、最小二乘法によってβの推定値をもとめると

$$b=\frac{\sum xy}{\sum x^2}$$

となるが、これを展開すれば

$$b=\frac{\sum x(\beta x+y''-\beta x'+\epsilon)}{\sum x^2}$$

$$=\beta+\frac{\sum xy''}{\sum x^2}-\beta\frac{\sum xx'}{\sum x^2}+\frac{\sum x\epsilon}{\sum x^2}$$

をうる。假設(i)(ii)(iii)より、右邊第二・第四項の期待値はゼロとなるから、

$$E(b)=\beta-\beta E\left(\frac{\sum xx'}{\sum x^2}\right)$$

したがって、この式の右邊第二項がゼロでないかぎり、bはβにくらべて偏りをもつことになる。

この偏りを處理するために、最近ベルクソンはきわめて巧妙な方法をしめした。[10] 彼によれば、多くの統計資料――とくに自然科學的な實驗資料――においては、x'よりもxが先決變數と考えられる。このようなばあいには、眞の値x'よりも觀察値xをコンスタントと考えた方がよい。そして、x'はむしろ觀察誤差x''の存在によって變動すると考え、前述の假設(ii)の代りに、

假設(iv)　$E(x_i''|x_1, x_2, \cdots\cdots, x_n)=0$　$(i=1, 2, \cdots\cdots, n)$

をもちいれば、

$$E\left(\frac{\sum xx''}{\sum x^2}\right)=0$$

となるから

$$E(b)=\beta$$

が成立する。すなわちbがβの不偏推定値となるが、そればかりでなく、x''・y''・εがそれぞれ正規分布をなすと考え、それぞれ系列相關を有しないとすれば、フィッシャー流の信賴度檢定も適用できることが證明されている。

以上のベルクソンの工夫は、いままでみのがされていた論點をたくみについている。しかし、經濟統計資料にこの種の誤差を想定するのはいかがなものであろうか。たとえば、つぎのような例が想起される。資本係數をひとつの經濟變量と考えて、その値を確定するのに國民資本の總額に關する統計資料を缺いているようなばあい、他國の實績からほぼ自國も4との想定をおいて、眞の値は4を中心に毎年動搖するとする。このさい、動搖のあり方を、4を中心とした對稱分布と前提すれば、假設(iv)が妥當する。だが、このような例はむしろまれで、大部分の經濟資料にかんする觀察誤差は、上述の假設(ii)と假設(iv)の中間に位置するような性質をもつものと思われる。

けれども、そのばあいも、ベルクソンの工夫にならって回歸係數の不偏推定値をもとめうるかなり有效な方法を考えうる。それは獨立變數のグルーピングをおこなうことである。上記、ベルクソンの工夫で、眞の値x'の變化の範圍をいくつかの區間に分割し、當該區間の中位數をもってx'を代表する。たとえば、x'_0とx'_0+hのあいだの代表値として$x'_0+\frac{1}{2}h$をもちいるというふうに。このときx'が誤差をふくまず、各區間においてx'がそれぞれの代表値を中心に對象的に分布しているものとすれば、グルーピングにともなう各區間の誤差はゼロの期待値をもって、假設(iv)を充すものと考えられる。それゆえ、

$$b=\frac{\sum xy}{\sum x^2}$$

はβの不偏推定値となるであろう。

ところで、われわれのとりあつかいうる資料は、誤差x''をともなった觀察値xである。いま、グルーピングがこれらの誤差によって影響をうけることがないと前提すれば、各グループの中位數は觀察誤差にかんして獨立になる。したがって、觀察誤差が眞の値x'と相關關係を有せず、その平均値がゼロであるとすれば、x''を觀察誤差と[8]グルーピングされた區間を中位數によって代表させることから生ずる誤差[9]との複合された誤差と考えることによって、假設(iv)を前提しうる。それゆえ、右の諸條件がげんみつに充されれば、回歸係數の不偏推定値をうることができるであろう。またたとえ、假設(iv)を規定しうるための諸條件がげんみつには充されなくても、グルーピングによって、すくなくともbの偏りを減少させることはできるものと思われる。[10] 上述の議論は、單に二變數のばあいばかりでなく、一般に多邊回歸のばあいにもそのまま擴張しうるものである。

(7) たとえば T. C. Koopmans, *Linear Regression Analysis of Economic Time Series*, 1937 を見よ。

(8) 時系列では、このほかに自己相關誤差の問題が生ずるが、ここでは差當りその問題にかんする假設を必要としない。

(9) J. Berkson, "Are There Two Regressions?" *Journal of American Statistical Association*, June 1950.

(10) 區間の代表値に平均値をとれば、グルーピング誤差がもたらす偏りはさけることができるが、變數の觀察誤差にもとづく偏りはさけることができない。

四

觀察誤差が假設(i)(ii)に規定されるような性質をもつばあい、回歸係數の推定値の偏りを消滅あるいは減少させる

手段として、補助變數 instrumental variable をもちいればよいことが、さいきんライアジールやゲアリーの研究によってしられている。[11] この節では、そのような工夫について考えてみよう。

獨立變數の眞の値 x' とは相關關係をもつけれども、誤差 x'' や ε とは獨立なある變數 z を想定する。このような z を補助變數とよぶ。いま、

$$b' = \frac{\sum zy}{\sum zx} = \frac{\sum z(\beta x' + \varepsilon)}{\sum z(x' + x'')}$$

とおけば

$$E(b') = \frac{\beta E(zx')}{E(zx')} = \beta$$

となるから、b' は β の不偏推定値たりうることになる。[12]

ところでこのばあい、一般に b' の分散 $V(b')$ は b の分散 $V(b)$ よりも大きいから、推定値として b' をもちいることはそれだけ精度がわるくなる。それがどのていどであるかをみるために、兩變數が觀察誤差をともなわないで測定されたばあい、すなわち $x=x'$, $y=y'$ なるときの相對精度を算定してみると、

$$V(b) = \frac{\sigma^2}{\sum x^2} \qquad \text{ただし } \sigma^2 = V(\varepsilon)$$

$$V(b') = \frac{\sigma^2 \sum z^2}{(\sum zx)^2}$$

から

$$\frac{V(b)}{V(b')} = \frac{(\sum zx)^2}{\sum z^2 \sum x^2} \leqq 1$$

をうる。

このように、補助變數をもちいることによって精度がわるくなるとすると、われわれはしいて補助變數をもちいて不偏推定値をえるよりも、むしろふつうの最小二乘法によって〈偏りをふくむ推定値〉に滿足した方がよいかもしれない。それではこの兩者の差は、どのようなひろがりをもつであろうか。

すでに指摘したように、$|E(b)|<\beta$, $|E(b')|=\beta$ であるから、$|b|<|b'|$ がえられる。そこで、前と同様に $x=x'$, $y=y'$ のときを考えると、

$$b-b'=\frac{\sum xy}{\sum x^2}-\frac{\sum zy}{\sum zx}$$
$$=\sum\left(\frac{x}{\sum x^2}-\frac{z}{\sum zx}\right)\varepsilon$$

したがって、

$$E(b-b')=0$$
$$V(b-b')=\sigma^2\sum\left(\frac{x}{\sum x^2}-\frac{z}{\sum zx}\right)^2$$
$$=\frac{\sigma^2}{\sum x^2}\left(\frac{V(b')}{V(b)}-1\right)$$

となる。すなわちbとb'の差の分散は、兩者の分散比に依存する。（ただし、兩變數の觀察誤差はゼロとして。）

さて、現實に補助變數法をもちいるばあいに問題となるのは、x'とは高度な相關關係をもちながら、x''とは獨立な變數をどうしてみつけるかということである。これにもさまざまな工夫を見出しうる。たとえばワルドがやったとつたえられる（筆者未見）ように、[13]

$$z=+1 \text{ or } 0 \text{ or } -1 \quad \text{for } x \gtreqless \text{ its median}$$

とするのも一法だろう。また、xを大いさにしたがって順位をつけて、その順位数をzと考えるのもよいだろう。いずれにしても、觀察誤差x''の値が大きくないのなら、有効な方法である。もしx''の値がかなり大きくて、xの順位に相當な影響を與えると考えられるならば、前節の方法にならってグルーピングをおこない、各區間ごとのxの平均値に順位をつければよい。[14]

(11) O. Reiersøl, "Confluence Analysis of Lag Moments and Other Methods of Confluence Analysis, *Econometrica*, Jan. 1941.
Ditto, *Confluence Analysis by means of Instrnmental Sets of Variables*, 1945.
Ditto, "Identifiability of Linear Relation between Variables Which are Subject to Error," *Econometrica*, July, 1950.
R. C. Geary, "Determination of Linear Relations between Systematic Parts of Varriables with Erros of Observation the Variances of Which are Unknown," *Econometrica*, Jan. 1949.

(12) βの信頼區間は、$y-\beta x$のzにたいする眞の回歸がゼロであることを利用して、つぎのように定めうる。いま、x''およびεが系列相關を有せず、平均値ゼロで一定の分散をもつものとすれば、

$$r^2=\frac{\{\sum(y-\beta x)z\}^2}{\sum z^2-\sum(y-\beta x)}$$

は、自由度$n-1$で、相關係數の平方のかたちをもつ。危険率5%のrの信頼限界をr_0とすれば

$$P(r^2\leqq r_0^2)=0.95$$

すなわち

$$P\left[\frac{(\sum yz)^2-2\beta\sum yz\sum zx+\beta^2(\sum xz)^2}{\sum z^2(\sum y^2-2\beta\sum xy+\beta^2\sum x^2)}\leqq r_0^2\right]=0.95$$

をうる。カッコのなかはβに關する二次不等式だから、これを因數分解して

$$P[(\beta-\beta_1)(\beta-\beta_1)\leqq 0]=0.95 \qquad (\beta_1<\beta_2)$$

をみちびくことができる。このばあいの$(\beta_1,\ \beta_2)$がβの95%の信頼區間である。

(13) A. Wald, The Fitting of Straight Lines if Both Variables are Subject to Error. *Annals of Mathematical Statistics*, No. 3, 1940

(14) ライアジールは前掲一九四一年の論文において、x'が系列相關を有しx''が系列相關をなさないばあいに、x'の lagged values を補助變數としている。

五

戰後、モデル・アナリシスの發達にともなって、經濟關係の計量にもいままでの單一方程式による最小二乘法的分析にかわり、いわゆる連立方程式接近法が採用されるようになったことはよく知られている。[15] さいごに、連立方程式接近法と補助變數法との關連についてかんたんにふれておこう。

y'とx'との回歸式と、x'とzとの回歸式とを連立させれば

$$\begin{cases} y'=\beta x'+\varepsilon \\ x'=\gamma z+\eta' \end{cases}$$

となるから

$$\begin{cases} y=\beta x+\xi \\ x=\gamma z+\eta \end{cases} \qquad \text{ただし} \begin{cases} \xi=\varepsilon-\beta x'' \\ \eta=\eta'+x'' \end{cases}$$

ここで、誤差項ξとηがzにかんして獨立であり、その平均値がゼロであるとすれば、γがゼロでないかぎり、誘導形法によるβの推定値として、

$$b=\frac{\sum xy}{\sum zx}$$

をうる。このばあい、εとbが同時正規分布をなすとすれば、b'がβの最尤推定値になることは前稿で證明したが、[16] 同時分布の性質に若干の限定をつければ、b'が不偏推定値になることもアンダーソンとラビンによってすでに證明されている。[17]

(15) T. C. Koopmans and Others. *Statistical Inference in Dynamic Economic Models*, 1950.
拙稿「經濟關係の計量とその推計學的基礎」經濟論叢　昭和二十五年九月

(16) 前掲拙稿一一三頁。

(17) T. W. Anderson and H. Rubin, "The Asymptotic Properties of Estimates of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics*, March, 1950.